When Worlds Collide: Quantum Probability From Observer Selection?

Robin Hanson *
Department of Economics
George Mason University†

August 9, 2001

Abstract

Deviations from exact decoherence make little difference for large-amplitude quantum states, but can make a big difference for small-amplitude states. This may provide a rationale for ignoring small-amplitude states when deriving the Born measurement probabilities from the many worlds interpretation. Interactions between large and small worlds may destroy the memory of observers in small amplitude worlds.

Introduction

Traditionally, quantum systems have been described as evolving according to two different rules. There is a usual deterministic linear evolution rule, which is occasionally replaced by a stochastic quantum measurement rule, which eliminates all but one diagonal element from the density matrix. This stochastic rule is ambiguous in several ways, however. Fortunately, recent research into "decoherence" has shown how off-diagonal elements are often naturally and dramatically suppressed via the coupling of a system to a large environment (**?**; **?**; **?**). This allows us to settle several measurement ambiguities, by assuming that the timing and observable of a quantum measurement coincide with the timing and observable of decoherence.

The many world interpretation of quantum mechanics seems to offer a way to settle several more measurement ambiguities (**?**; **?**). A problem with the many worlds approach, however, is that to predict the standard Born rule for measurement probabilities, it appears one must assume the irrelevance of worlds whose magnitude is relatively very small. Yet the vast majority of worlds are this small, and it is not clear why one should discount the experiences of observers in small worlds (**?**; **?**). After all, if worlds split according to standard ideal measurement processes, the future evolution of each world should be independent of the evolution of other worlds.

*For their comments, I thank Nick Bostrom and David Deutsch.

†*rhanson@gmu.edu* http://hanson.gmu.edu 704-993-2326 FAX: 704-993-2323 MSN 1D3, Carow Hall, Fairfax VA 22030

This paper suggests that we can reconcile the many worlds approach with the Born rule if we realize that decoherence is never exact, and so worlds do not exactly split according to an ideal measurement process. The evolution of the density matrix describing a world is influenced both by internal autonomous dynamics, *and* by cross-world influences from off-diagonal density matrix terms. While decoherence makes these off-diagonal terms small relative to a large world, they can be large relative to a small enough world. So while to a good approximation large worlds evolve autonomously, the evolution of small worlds can be dominated by the influences from larger worlds.

This may plausibly either destroy the observers in small worlds, or change then into observers who remember the large world events. In either case, most observers in all worlds would remember having observed frequencies near that predicted by the Born rule, even if in fact most worlds do not have such frequencies. A "mangled worlds" variation on the many worlds interpretation may thus predict the Born rule for quantum probabilities.

This paper will first review the basics of quantum measurement, decoherence, and the many worlds interpretation, and then discuss the implications of inexact decoherence for the autonomy of small world evolution.

Quantum Measurement

Quantum mechanics has traditionally described systems by unit-magnitude Hilbert space vectors $|\,\psi\,\rangle$ which evolve according to two different rules. Usually, vectors evolve deterministically according to the standard linear rule

$$i\hbar\frac{d}{dt}|\,\psi\,\rangle = H|\,\psi\,\rangle,$$

but occasionally they evolve non-deterministically according to

$$|\,\psi\,\rangle = \sum_a \,|\,a\,\rangle\langle\,a\,|\,\psi\,\rangle \text{ becomes } |\,a\,\rangle \text{ with probability } |\langle\,a\,|\,\psi\,\rangle|^2.$$

This second process is said to correspond to the measurement of a value a for some "observable," where the $|\,a\,\rangle$ and a are respectively eigenvectors and eigenvalues of that observable's Hermitian operator A. More precisely, given a complete set of orthogonal projection operators $\{P_a\}_a$ (so $\sum_a P_a = 1$ and $P_aP_b = \delta_{ab}P_a$), a measurement on $|\,\psi\,\rangle$ produces a normalized $P_a|\,\psi\,\rangle$ with probability equal to the magnitude-squared of $P_a|\,\psi\,\rangle$.

In place of a Hilbert space vector $|\,\psi\,\rangle$, one can equivalently describe a quantum system in terms of a unit-trace Hermitian "density matrix" ρ, which usually evolves according to

$$i\hbar\frac{d}{dt}\rho = H\rho - \rho H, \tag{1}$$

but occasionally evolves according to

$$\rho \text{ becomes a normalized } P_a\rho P_a \text{ with probability } \mathrm{tr}(P_a\rho), \tag{2}$$

where again the outcome $P_a\rho P_a$ coincides with the observation of a value a for observable $A = \sum_a aP_a$.

People have for seventy years wrestled with ambiguities in the above descriptions of quantum dynamics. What determines the time t and projections $\{P_a\}_a$ for each measurement? And to what extent do $|\,\psi\,\rangle$ and ρ describe the system itself, as opposed to our knowledge of the system?

Decoherence

Great progress has recently been made regarding the ambiguities of time and observable. In the density matrix formulation of quantum dynamics, a key distinguishing feature of the measurement process described in equation 2 is that it displays "decoherence." That is, measurement eliminates the off-diagonal elements of ρ in the A representation of ρ. Recently, detailed analyses of many specific physical systems have shown that such decoherence can result from the interaction of a system with a large environment via standard linear quantum evolution.

The usual decoherence scenario describes a total system $|\,\psi\,\rangle|\,e\,\rangle$, which is a particular quantum system $|\,\psi\,\rangle$ coupled to a large environment $|\,e\,\rangle$. This scenario considers not the total density matrix ρ_T, but only the part of that matrix describing only the system $|\,\psi\,\rangle$. That is, while linear quantum evolution must preserve the off-diagonal elements of the total density matrix ρ_T, it need not do so for the system density matrix

$$\rho = \mathrm{tr}_E(\rho_T),$$

where tr_E denotes a trace across the environment subspace. Detailed analyses of many specific situations have shown that initial system coherence in ρ is usually transfered very quickly to the environment. That is, the physics of a particular situation often chooses particular projections $\{P_a\}$ for the system and then induces a measurement-like evolution

$$\rho = \sum_{ab} \rho_{ab} \quad \rightarrow \quad \approx \quad \rho' = \sum_a \rho'_{aa}, \tag{3}$$

where $\rho_{ab} = P_a \rho P_b$. It seems that once coherence has been transfered from a quantum system to a large environment, it becomes for all practical purposes impossible to measure. If so, then we can eliminate many ambiguities in measurement theory by assuming that the measurement process of equation 2 only happens after decoherence of the same projections P_a as in equation 3.

Of course there still remains the ambiguity of the extent to which state vectors describe reality, as opposed to our knowledge of reality. And the measurement process of equation 2 is still distinguished from the decoherence process of equation 3 by the fact that while in the $\{P_a\}_a$ representation decoherence eliminates the off-diagonal elements of ρ, measurement also eliminates all but one of the diagonal elements of ρ. And it does this non-deterministically.

Many Worlds Interpretation

Forty years ago Hugh Everett proposed the "many worlds," or "relative state," interpretation of quantum mechanics as a resolution of these remaining ambiguities (**?**; **?**). (Many worlds actually came first, and much of the later decoherence research was inspired by attempts to address ambiguities that the many worlds interpretation did not directly address.) The many worlds interpretation posited that vectors $|\,\psi\,\rangle$ are literally real, and that they describe all physical systems, including human observers. It posited furthermore that systems only evolve deterministically according to the standard linear quantum rule.

Before an ideal many-worlds measurement, a total system vector is a product of an observed system vector $|\,\psi\,\rangle$ and an observer system vector $|\,O\,\rangle$. During an ideal measurement, the total system then evolves deterministically according to

$$|\,\psi\,\rangle|\,O\,\rangle = \left(\sum_a |\,a\,\rangle\langle\,a\,|\,\psi\,\rangle\right)|\,O\,\rangle \rightarrow \sum_a \langle\,a\,|\,\psi\,\rangle|\,a\,\rangle|\,O_a\,\rangle, \tag{4}$$

where $|\,O_a\,\rangle$ describes an observer who has observed the value a. The apparent non-determinism of measurement evolution is resolved by positing that each $|\,a\,\rangle > |\,O_a\,\rangle$ describes a different "world," where a different copy of the original observer has measured a different value of a.

The many worlds interpretation in essence posits that a careful analysis of standard linear quantum evolution will show that it reproduces all of the phenomena usually explained by invoking non-deterministic quantum evolution. While the many worlds interpretation did not originally say when and along what projection basis worlds split, the recent decoherence analyses have gone a long way to offering answers to such questions. That is, it seems plausible to say that the world of an observer splits when that observer observes a system displaying quantum interference effects, and when this observer and observed are coupled to a large environment so as to produce decoherence across multiple relevant states of the observed quantum system.

There remains, however, one important measurement phenomena that the many worlds interpretation has not yet adequately accounted for. That phenomena is the "Born rule," i.e., the particular probability distribution with which measurement produces states $|\,a\,\rangle$. The Born rule is that vector $|\,a\,\rangle$ is seen with probability $|\langle\,a\,|\,\psi\,\rangle|^2$, or that a normalized ρ_{aa} is seen with probability $\mathrm{tr}(P_a\rho)$. While this distribution can be derived from assumptions of unitary symmetry (**?**), a straightforward many worlds observer-selection analysis gives a different answer (**?**; **?**).[1] If we simply count the number of worlds in which observers see different long-run measurement frequencies, we find that the vast majority of such worlds will display long-run frequencies consistent with a uniform distribution. Under a uniform distribution, a normalized ρ_{aa} is seen with probability $1/N$, where N is the number of projections in the set $\{P_a\}_a$.

Everett tried to address this problem by showing that worlds $|\,s\,\rangle$ containing long-run frequencies which differ substantially from the Born frequencies have a much smaller amplitude than worlds $|\,L\,\rangle$ containing frequencies very near the Born rule. That is, $\mathrm{tr}(\rho_{ss}) \ll \mathrm{tr}(\rho_{LL})$. In the limit of an infinite number of non-trivial measurements, the amplitude of non-Born frequency worlds approaches zero, i.e., $\mathrm{tr}(\rho_{ss})/\mathrm{tr}(\rho_{LL}) \rightarrow 0$. It is not clear, however, that there are ever an infinite number of non-trivial measurements, or that observers should be bothered by living in small but non-zero amplitude worlds. After all, if world splittings hap-

[1] One can reconcile the many worlds interpretation with the Born rule by postulating an infinite number of worlds correspond to each possible measured state, which then split during a measurement in proportion to the Born rule measure (**?**). This approach seems ad hoc, however, and leaves open the question of how this splitting physically occurs.

pen according to the ideal measurement process described in equation 4, the future evolution of a world should be entirely independent of the amplitude of that world.

Inexact Decoherence

In fact, however, real measurements are not usually exactly ideal. And if equation 3 is only approximate then equation 4 is not exactly right either. The purpose of this paper is to suggest that this discrepancy may allow us to predict the Born rule from within the many worlds interpretation.

The basic idea is that deviations from equation 3 allow interactions, or "collisions," between large worlds $| L \rangle$ and small worlds $| s \rangle$, interactions which may be strong enough to typically erase the measurement records in very small worlds $| s \rangle$. If so, then observers who remain with intact measurement records will primarily be in large worlds $| L \rangle$ which satisfy the Born rule. In support of this suggestion, let us make a rough estimate of the relative magnitudes of the processes which influence the evolution of a single large world and single small world.

Consider first the relative size of the small world, given by $\delta(t)$ in

$$|\rho_{ss}| \approx \delta(t)^2 |\rho_{LL}|, \tag{5}$$

where magnitude is $|\rho_{aa}| = \mathrm{tr}(\rho_{aa})$. Everett showed that $\delta(t)$ should fall as the (square root of the) ratio of the likelihoods the Born rule would assign to the frequencies observed in the two worlds. For a sequence of identical measurements with substantial uncertainty, for example, $\delta(t)$ should fall exponentially with the number of measurements included the frequency.

Next consider the relative magnitude of diagonal and off-diagonal terms, ρ_{ab} for $a \neq b$. Roughly following (**?**), we may describe a degree of decoherence ϵ as[2]

$$|\rho_{ab}|^2 \leq \epsilon^2(t) |\rho_{aa}||\rho_{bb}|,$$

where for fine grain projections $P_a = | a \rangle\langle a |$, we can define magnitude as $|\rho_{ab}| = \langle a | \rho_{ab} | b \rangle$ (how should it be defined for coarser projections?).Let us therefore write

$$|\rho_{Ls}|^2 \approx |\rho_{sL}|^2 \approx \epsilon^2(t) |\rho_{LL}||\rho_{ss}|.$$

If we combine this equation with equation 5 we obtain these relative magnitudes

$$\begin{array}{ccccc} \rho_{LL} & & & \approx & 1 \\ \rho_{Ls} & \approx & \rho_{sL} & \approx & \epsilon\delta \\ \rho_{ss} & & & \approx & \delta^2 \end{array}$$

The key point here is that even when coherence ϵ is very small, if the Born likelihood ratio δ is smaller still, then off-diagonal terms like ρ_{Ls} can still have larger magnitudes than small world diagonal terms like ρ_{ss}.

[2] It would also work to substitute $(|\rho_{aa}| + |\rho_{bb}|)^2$ for $|\rho_{aa}||\rho_{bb}|$ in the definition of ϵ.

Can the likelihood ratio δ really be smaller than the coherence ϵ? Detailed analyses of many specific situations has found that while the coherence $\epsilon(t)$ typically falls at a very rapid exponential rate for substantial periods, it eventually asymptotes to a small but non-zero level. The happens, for example, in several models where a particular observable of a single simple particle is continually "measured" by an infinite (or finite) environment, and where this measurement continues on forever (**?**; **?**; **?**; **?**). The likelihood ratio, in contrast, should continue to get smaller without limit. If a similar behavior is found in models of environmental measurement of non-commuting observables, that would support the following conjecture.

Conjecture 1 *For worlds with long-run frequencies substantially different from the Born rule, the Born likelihood ratio $\delta(t)$ eventually falls and stays well below the coherence $\epsilon(t)$.*

Worlds Colliding

Let us now consider the co-evolution of a single Born-rule large world $|\, L\,\rangle$ and a single small other-frequency world $|\, s\,\rangle$, which have almost but not exactly decohered due to coupling with some environment. The total system of worlds plus environment must evolve according to equation 1, but the partial system of the worlds alone need not. If we assume the interaction with the environment is weak, however, we can write

$$i\hbar\frac{d}{dt}\rho = H\rho - \rho H + S$$

where H describes the evolution of isolated worlds, and S describes the change in ρ due to interaction with the environment. If we decompose this according to the two projections P_L and P_s for the two worlds, we get

$$i\hbar\frac{d}{dt}\rho_{LL} = \quad H_{LL}\,\rho_{LL} - \rho_{LL}\,H_{LL} \quad +(H_{Ls}\,\rho_{sL} - \rho_{Ls}\,H_{sL}) + S_{LL} \tag{6}$$

$$i\hbar\frac{d}{dt}\rho_{ss} = \quad H_{ss}\,\rho_{ss} - \rho_{ss}\,H_{ss} \quad +(H_{Ls}\,\rho_{sL} - \rho_{Ls}\,H_{sL}) + S_{ss} \tag{7}$$

$$i\hbar\frac{d}{dt}\rho_{Ls} = \quad H_{LL}\,\rho_{Ls} - \rho_{LL}\,H_{Ls} \quad +H_{Ls}\,\rho_{ss} - \rho_{Ls}\,H_{ss} + S_{Ls} \tag{8}$$

$$i\hbar\frac{d}{dt}\rho_{sL} = \quad H_{sL}\,\rho_{LL} - \rho_{sL}\,H_{LL} \quad +H_{ss}\,\rho_{sL} - \rho_{ss}\,H_{sL} + S_{sL} \tag{9}$$

Equations 6 and 7 describe the evolution of the two worlds, and the terms in parentheses describe evolution due to influence from off-diagonal terms. Equations 8 and 9 describe the evolution of the off-diagonal terms, in part due to influence from the two worlds.

If the various H terms are of similar magnitudes, then the autonomy of a world's evolution depends primarily on the relative magnitudes of the various density matrix terms. We have assumed that $|\rho_{ss}| \approx \delta^2|\rho_{LL}|$ and $|\rho_{sL}| \approx |\rho_{Ls}| \approx \epsilon\delta|\rho_{LL}|$. This implies that (for S small) ρ_{LL} is by far the largest influence on the evolution of ρ_{LL}. That is, the large world evolves autonomously to a very good approximation. In contrast, for $\delta < \epsilon$ the evolution of ρ_{ss} is determined at least as much by ρ_{sL} and ρ_{Ls} as by ρ_{ss}. And the evolution of ρ_{sL} and ρ_{Ls} is dominated by ρ_{LL}. That is, the evolution of a small enough world is mostly slaved to the evolution of the large world, via the off-diagonal intermediaries.

If, as we have assumed, P_s and P_L are the projections that make the two worlds seem the most decoherent, then it seems that we simply cannot consider the small world s to be evolving autonomously as suggested by the idealized many worlds measurement of equation 4. This suggests that the evolution of typical measurement records and observers in small world s will be determined primarily by influences from measurement records and observers in the large world L. It seems plausible that such strong influence may either destroy such small world measurement records and observers, since as physical systems they were designed to evolve under a different Hamiltonian, or this strong influence may change those measurement records and observers into ones much like those found in the large world. This suggests the following conjecture

Conjecture 2 *When coherence ϵ is large enough relative to likelihood ratio δ, small world human observers will typically either fail to exist or will remember the measurement frequency of large worlds.*

Taken together, conjectures 1 and 2 suggest when a large Born-rule world interacts with a small other world, the human observers in both worlds will remember having observed near Born frequencies for quantum measurements. This suggests further that when a small number of large Born-rule worlds interact with a large number of small worlds of other frequencies, most human observers in all the worlds will remember having observed near Born quantum measurements frequencies. This can be true even if in the vast majority of worlds, by count, the actual measurement frequency is closer to a uniform distribution.

Note that conjecture 2 need only apply to typical current human observers. It could be that, by using quantum error correction codes, quantum computers and the humans that observe them will be better able to resist the influence of larger worlds via off diagonal terms. If so, perhaps such observers will be able to remember non-Born measurement frequencies.

Conclusion

This paper has suggested that the many worlds interpretation can predict the Born rule or quantum measurement probabilities if two conjectures hold. First, although the coherence between different worlds typically falls very rapidly, it eventually falls slowly, even given repeated measurements of non-commuting observables. Second, although coherence may be very small, when it is large enough compared to the relative magnitude of worlds, typical human observers in small worlds are either destroyed, or fail to remember the measurement frequencies of those small worlds. Both of these conjectures should be open to confirmation or rejection by more detailed analysis of the evolution of specific quantum systems.

If these conjectures are confirmed, then a "mangled worlds" variation on the many worlds interpretation seems able to predict that the observers in the vast majority of worlds will recall near Born frequencies for quantum measurements, even if in fact most such worlds contain very different frequencies. The many worlds interpretation would then be one step closer to a satisfactory account of many ambiguities of quantum measurement.